




Assessing the methodological quality and risk of bias of systematic reviews: primer for authors of overviews of systematic reviews

Carole Lunny ^{1,2} Salmaan Kanji,^{3,4} Pierre Thabet,⁵ Anna-Bettina Haidich,⁶ Konstantinos I Bougioukas,⁶ Dawid Pieper⁷

For numbered affiliations see end of article.

Correspondence to: Dr Carole Lunny, Knowledge Translation Program, Li Ka Shing Knowledge Institute, St Michael's Hospital, Toronto, Canada; carole.lunny@ubc.ca

Cite this as: *BMJMED* 2024;3:e000604. doi:10.1136/bmjmed-2023-000604

Received: 24 April 2023
Accepted: 3 May 2024

Carole Lunny and colleagues consider methods such as AMSTAR-2 and ROBIS tools to evaluate the methodological quality and risk of bias of systematic reviews of intervention effects that are included in overviews of reviews

Introduction

Overviews of reviews, which synthesise the findings of systematic reviews,¹ have significantly increased in publication over the past decade.² However, the terminology used to describe them is not agreed in consensus, with terms such as umbrella reviews, meta-reviews, and reviews of reviews being used interchangeably to mean overviews of reviews. Methods research has been ongoing since the 2010s to develop effective approaches for conducting overviews of reviews and addressing their unique characteristics.^{3–7} Overview authors use various approaches to assess the methodological quality and risk of bias in their included systematic reviews, and they apply these assessments to inform the overviews' results and conclusions. However, proper use of tools for this purpose require training, time, and an appreciation of their strengths and limitations. This methods primer aims to address the inconsistency in assessing and reporting bias in systematic reviews of intervention effects included within overviews, and focuses on presenting the different validated tools,

comparing them, and providing guidance on the interpretation and reporting of these assessments.

Assessing the methodological quality and risk of bias Assessment tools

Many tools exist to evaluate methodological quality and risks of bias in systematic reviews, but they have been developed with different purposes, and choosing among them is difficult. More than 40 critical appraisal tools exist to evaluate the content and measurement properties of systematic reviews.^{8–9} After these reviews were published, two new tools were developed (ie, ROBIS and AMSTAR-2^{10–11}), and one is under development (Risk of Bias in Network Meta-Analysis (RoB NMA)^{12–13}).

In 2016, ROBIS was developed to assess risk of bias in systematic reviews,¹¹ ROBIS consists of three phases: assessment of relevance (optional), identification of bias concerns with the review process, and judgement of the overall risk of bias in the review. The tool focuses on four domains: study eligibility criteria, identification and selection of studies, data collection and study appraisal, and synthesis and findings. ROBIS helps reviewers identify potential biases in these domains by asking specific questions related to the review's methods and reporting. The tool underwent content validity and reliability testing to ensure its accuracy and consistency in assessing the risk of bias in systematic reviews.

In 2017, an update to AMSTAR, called AMSTAR-2,¹⁰ aimed to assess methodological quality of systematic reviews, and involved inter-rater reliability and usability testing. AMSTAR-2 consists of 16 items that evaluate various aspects of the systematic review process, including the research question formulation, study selection and data extraction, assessment of risk of bias in individual studies, consideration of publication bias, and appropriate statistical analysis. This tool also assesses the overall methodological quality and risk of bias in the review, providing a comprehensive evaluation.

The decision about how to evaluate overall risk of bias for ROBIS is made at the assessors' discretion, as opposed to the AMSTAR-2 overall judgement, which is prescribed by AMSTAR-2 guidance. Examples of how to interpret methodological quality and risk of bias assessments, and how to make an overall judgement are found in [box 1](#).

The AMSTAR-2 and ROBIS tools were designed to assess systematic reviews with pairwise

KEY MESSAGES

- ⇒ Systematic reviews underpin evidence based healthcare decision making, but flaws in their conduct may lead to biased estimates of intervention effects and hence invalid recommendations
- ⇒ Overviews of reviews (also known as umbrella reviews, meta-reviews, or reviews of reviews) evaluate biases at the systematic review level, among others, but proper use of tools for this purpose require training, time, and an appreciation of their strengths and limitations
- ⇒ AMSTAR-2 and ROBIS are the two most popular and rigorous critical appraisal tools used for appraising systematic reviews
- ⇒ The AMSTAR-2 16-item checklist focuses on methodological quality of systematic reviews of healthcare interventions, and incorporates aspects of review conduct, reporting comprehensiveness, and risk of bias as specific items
- ⇒ ROBIS is a domain based tool with 19 items focusing on risk of biases in a systematic review (eg, selective reporting of outcomes or analyses) of healthcare interventions and contains items related to risk of bias in results and conclusions, relevance, and an item about risk of interpretation bias or 'spin'

BOX 1 | DECISION RULES: HOW TO DECIDE THAT THE RESULTS OF A REVIEW ARE OF HIGH QUALITY OR AT LOW RISK OF BIAS OVERALL

Decision rules are a priori strategies used to specify rules to define explicitly how each item is rated, as well as how an overall judgement is made about a specific systematic review with the AMSTAR-2 and ROBIS tools. In the case of AMSTAR-2, the authors who are using the tool stipulate how to come to an overall high quality rating in the results of the review, but not how to rate each item. For example, item 15 of AMSTAR-2 asks assessors whether an adequate investigation of publication bias (small study bias) was conducted and whether its likely effect on the results was discussed. However, the AMSTAR-2 team did not specify what happens when 10 studies or fewer were included (ie, the analysis will be underpowered to detect publication bias), what methods to detect publication bias are recommended, and if publication bias is detected, how it should be discussed (ie, as a systematic review limitation).

The ROBIS tool equally does not specify what decision rules should be used for assessment of risk of bias, nor how to come to an overall judgement. For example, item 4.6 of ROBIS ("Were biases in primary studies minimal or addressed in the synthesis?") is similar to item 12 of AMSTAR-2 ("If meta-analysis was performed, did the review authors assess the potential impact of risk of bias in individual studies on the results of the meta-analysis?"). Of note, risk of bias should be assessed in any systematic review regardless of whether a meta-analysis was performed. A possible decision rule for answering these two questions when considering whether bias was addressed and considered in the results and their interpretation could be to respond "Yes" or "Probably/Partial Yes" if:

- ⇒ All studies received a low risk of bias rating; and
- ⇒ Studies were judged at high risk of bias and sensitivity analyses (grouping high v low risk studies in a meta-analysis) or adjustment approaches were used

For a "No" response:

- ⇒ Important biases were suspected to have been in the included studies that have been ignored by the review authors; or
- ⇒ Risk of bias was not assessed at all in the included studies; or
- ⇒ Bias was assessed but authors did not incorporate it into findings, discussion, and conclusions

Based on the above decision rules, how would the following statement be rated? "We planned on conducting sensitivity analysis on the studies based on their level of risk of bias. Most of the included studies had a similar risk of bias across all the domains except for industry sponsorship bias and incomplete data for total testosterone. Due to the inadequate number of studies, we were not able to conduct a sensitivity analysis on the included studies based on industry sponsorship."

For overall judgements, a decision rule could be that if one or more ROBIS domains are at high risk of bias, then the overall study is deemed at high risk of bias. For AMSTAR-2, the authors of the tool have stipulated that the review is considered of low or critical low quality when any of the subset of seven 'critical' items have one or more critical flaws. While the decisions about how to rate the items and make overall judgements can be debated, the grounds on which overview authors make these decisions should be noted explicitly in the manuscript or in an appendix, as then the assessment results will be transparent and reproducible.

- ⇒ Cautionary note: empirical evidence does not currently support the assignment of scores to items that are met in a risk of bias tool followed by the summation or averaging of these scores to produce a numerical measure of risk of bias. A thoughtful, nuanced, and customised overall judgement is required that considers all items with suspected bias on the basis of specific context.

meta-analysis only. A more recent tool under development aims to assess the potential biases and limitations in network meta-analyses.^{12 13} Guidance documents (eg, Cochrane¹⁴ and JBI¹⁵) recommend overview authors use ROBIS or AMSTAR-2 when comparing and critically appraising systematic reviews over other available tools. Figure 1 presents two example assessments conducted by our team, the ROBIS assessment of Normansell and colleagues¹⁶ is presented at the domain level, and the AMSTAR-2 assessment of Puig and colleagues¹⁷ is presented by item. Items are backed by quotes and rationales to

support the answers chosen, for full transparency, and to help when comparing assessments between two independent assessors (figure 2).

Comparison of AMSTAR-2 and ROBIS

Both the AMSTAR-2 and ROBIS tools provide structured guidelines for reviewers to evaluate and report on methodological strengths and weaknesses as well as potential biases in systematic reviews, contributing to the overall reliability and credibility of the evidence presented. Considerable overlap exists between the items of the two tools (figure 1). In the

Phase 2				Phase 3			
First author, Year	Summarise concerns identified in each domain:				Where conclusions supported by evidence:		
	Domain 1: Concerns regarding specification of study eligibility criteria (low, high, unclear)	Domain 2: Concerns regarding methods used to identify and/or select studies (low, high, unclear)	Domain 3: Concerns regarding methods used to collect data and appraise studies (low, high, unclear)	Domain 4: Concerns regarding the synthesis and findings (low, high, unclear)	A. Did the interpretation of findings address all of the concerns identified in domains 1 to 4?	B. Was the relevance of identified studies to reviewers' research question appropriately considered?	C. Did the reviewers avoid emphasising results on their statistical significance?
Normansell, 2018	Low	Low	Low	Low	Yes	Yes	Yes
Quotes and rationale	<ul style="list-style-type: none">• No access to the protocol, but the section "deviations from the protocol" is transparent• PICO well defined• No restrictions: "... with no restriction on language of publication"• At least two review authors screened the search results for eligible studies"	<ul style="list-style-type: none">• "We searched the Cochrane Airways Trials Register, which contains records compiled from multiple electronic and handsearched resources"• "We also searched trial registries and reference lists of primary studies." Relevant manufacturer websites as well• Full strategy in appendix 1 and 2• No restrictions, including language• "At least two review authors screened the search results for eligible studies"	<ul style="list-style-type: none">• "We extracted outcome data, assessed risk of bias in duplicate, and resolved discrepancies by involving another review author"• Detailed tables were provided	<ul style="list-style-type: none">• All studies accounted for• Good description of deviations in the section "Deviations from protocol"• "We undertook meta-analyses only when this was meaningful, that is, when treatments, participants, and the underlying clinical questions were similar enough for pooling to make sense"• They did subgroup and sensitivity analyses• "We were able to perform only very limited subgroup and sensitivity analyses owing to the small number of trials identified, and we found no evidence of important effect modification..."• "Finally, we had concerns about risk of bias related to lack of blinding in outcomes"• "Our confidence was reduced for all outcomes by suspicion of publication bias: despite the availability of a common treatment for asthma, we identified only six eligible RCTs, suggesting that unpublished data may exist. However, we were not able to formally explore this by using a funnel plot because we identified an insufficient number of studies..."	<ul style="list-style-type: none">• Authors provided a detailed list of the limitations of the review	<ul style="list-style-type: none">• The review question matched the PICO of the included studies	<ul style="list-style-type: none">• "We found very limited evidence that the antibiotics may help people having asthma attacks"• They qualified their findings with lots of caveats about the limitations of the data
					Low risk of bias		

Figure 1 | Example assessments using ROBIS of Normansell¹⁶ and AMSTAR-2 of Puig¹⁷. The ROBIS assessment is presented by domain and the AMSTAR-2 assessment by individual items. ROBIS's phase one, where the assessor considers the relevance of the systematic review questions to the overview's question, is not shown. The decision about how to evaluate overall risk of bias for ROBIS is made at the assessors' discretion, as opposed to the AMSTAR-2 overall judgement, which is prescribed by AMSTAR-2 guidance

Author, Year	Item 1: PICO components	Item 2: Protocol	Item 3: Study design rationale	Item 4: Comprehensive search strategies	Item 5: Duplicate study screening	Item 6: Duplicate data extraction	Item 7: List of excluded studies	Item 8: Characteristics of studies (tables)	Item 9a: Risk of bias assessment
Puig, 2017	Yes	Yes	No	Yes	Yes	Yes	No	Yes	No
Quotes and rationale	<ul style="list-style-type: none"> Published articles were included if they investigated one of the following treatments: all DMARDs including biologics used for psoriasis treatment. The comparisons had to be done either v a placebo or an active treatment 	<ul style="list-style-type: none"> The selection of databases, eligibility criteria, outcomes of the review and analyses methods were defined a priori in an internal protocol 	<ul style="list-style-type: none"> No rationale: "Study design included RCTs, non-RCTs, and observational studies" 	<ul style="list-style-type: none"> Medline (including Medline-in-Process), PubMed, Embase, the Cochrane Central Register of Controlled Trials, and National Health Service Economic Evaluation Database" More than two databases searched, grey literature searched through Cochrane Centre Register of Trials 	<ul style="list-style-type: none"> Study selection, data extraction were performed independently by two assessors. Any discrepancy was reconciled by a third independent reviewer" 	<ul style="list-style-type: none"> See item 5 	<ul style="list-style-type: none"> Authors did not provide a list of excluded studies with justification 	<ul style="list-style-type: none"> Table 2: The demographics and disease characteristics of study participants at baseline" 	<ul style="list-style-type: none"> Jadad for RCT was used, which is not an appropriate tool
Author, Year	Item 9b: PICO components	Item 10: Funding sources from included studies	Item 11: Appropriate methods for statistical combination of results	Item 12: Impact of bias assessment results in meta-analysis	Item 13: Used risk of bias in interpreting results	Item 14: Explanation of heterogeneity	Item 15: Publication bias	Item 16: Conflict of interests	Overall judgement
Puig, 2017	No	No	Yes	No	No	Yes	Not applicable	Yes	Critically low quality
Quotes and rationale	<ul style="list-style-type: none"> Downs and Black's for non-RCT was used, which is not an appropriate tool 	<ul style="list-style-type: none"> For the four studies identified from systematic search, funding was not reported: the three additional trials included were all funded by Novartis 	<ul style="list-style-type: none"> "Classical statistically significant differences between DLQI effect in PASI 75-89 and PASI 90 responders, at 0.05 level, were assessed using one sided Bayesian P values; these are the Bayesian probability that DLQI effect was greater in one responder group than the other. Meta-analysis of studies was conducted using OpenBUGS (version 3.2.2 rev 1063)" 	<ul style="list-style-type: none"> Risk of bias was assessed but authors did not incorporate it into findings and conclusions 	<ul style="list-style-type: none"> Risk of bias was not used in the interpretation of the results 	<ul style="list-style-type: none"> See item 11. Bayesian analysis accounted for heterogeneity 	<ul style="list-style-type: none"> Fewer than ten studies were included which made subgroup analysis and funnel plot analysis not possible 	<ul style="list-style-type: none"> Sponsored by Novartis Pharma AG, Basel, Switzerland 	<ul style="list-style-type: none"> More than one critical flaw (item 2, 7, 9, 13) therefore the review is deemed critically low quality

Critical flaw

Note: more than one critical flaw (ie, a "no" response) and the review is deemed critically low quality

Figure 2 | PICO framework stands for patient or problem, intervention or exposure, comparison or control, and outcomes. DLQI=dermatology life quality index; DMARDs=disease modifying anti-rheumatic drugs; PASI=psoriasis area and severity index; RCT=randomised controlled trial

Table 1 | Comparison of the AMSTAR-2 and ROBIS tools

Name of tool	Concept measured	Areas of research	Number of items	Items unique to tool (and associated item number)	Type of overall ratings	Time (range across three studies*)
AMSTAR-2	Methodological quality	Intervention	16 items	<ul style="list-style-type: none">Whether the authors explained why they selected the study designs for inclusion in the review (item 4)Whether authors provided a list of excluded studies and justification for exclusions (item 7)Whether authors reported on sources of funding for included studies (item 10)Whether authors reported their own potential competing interests (item 16)	Overall confidence in the results of the review (critically low, low, moderate, or high), based on assessments of critical domains, predetermined by user (appraisers can alternatively use the critical domains suggested by the AMSTAR-2 developers)	12-18 min
ROBIS	Risk of bias	Intervention, diagnosis, prognosis, and biological cause	19 items organised into four domains	<ul style="list-style-type: none">Applicability of the review question to the assessors' research question (ie, do they match) (phase 1)Whether eligibility criteria were appropriate for research question (item 1.2)Whether restrictions in eligibility criteria based on study characteristics (eg, date, sample size, study quality, outcomes measured) were appropriate (item 1.4)Whether restrictions on eligibility criteria (eg, publication status or format, language, availability of data) were appropriate (item 1.5)Whether search restrictions based on date, publication format, or language were appropriate (item 2.4)Whether all relevant study results were collected for use in the synthesis (item 3.3)Whether studies were selectively missing from the synthesis (item 4.1)Whether all predefined analyses were reported (item 4.2)Whether authors acknowledged the limitations (and biases) when interpreting the results of their review (phase 3)Whether authors avoided emphasising results on the basis of their statistical significance? (phase 3)	Risk of bias ratings assigned for each domain and overall (low, high, or unclear)	16-24 min

*Perry et al¹⁸ had a slightly lower time for AMSTAR-2 than for ROBIS (14 min v 24 min); Pieper et al²¹ had a mean time for scoring AMSTAR-2 slightly higher than that for ROBIS (18 min v 16 min); and Gates²⁰ had lower time for AMSTAR-2 than for ROBIS (12 min v 17 min).

documentation for each tool, AMSTAR-2 states that it was developed for systematic reviews of healthcare interventions whereas ROBIS states that it is aimed at reviews of healthcare interventions, diagnosis, prognosis, and biological cause. In practice, the ROBIS tool is generic and its signalling questions relate to interventions in the clinical or public health fields. Questions specific to systematic reviews of diagnosis, prognosis, and biological cause are not found in the tool. AMSTAR-2 was developed to assess methodological quality (which includes indicators of risk of bias) while ROBIS was developed primarily to assess risk of bias but also includes items that address methodological quality.

AMSTAR-2 focuses more on reporting comprehensiveness (eg, reporting of study designs for inclusion and reporting on excluded studies with justification) and methodological quality or transparency constructs (eg, pre-established protocol, sources of funding of primary studies, and reviewers' competing interests). Whereas ROBIS focuses on items related to identification of the different biases (eg, selective reporting of outcomes or analyses and publication bias). Bias occurs when factors systematically affect the results and conclusions of a review and cause them to be systematically different from the truth.¹ Systematic reviews affected by bias can be inaccurate; for example, finding false positive or false negative intervention effects by systematically over or under estimating the true effect in the target population. Methodological quality focuses on methodological features associated with internal validity. In theory, assessing risk of bias is the preferred approach because a review might have good methodological quality while still being at high risk of bias. For example, a systematic review might have been conducted according to stated guidance, but some relevant databases were not searched for evidence (database selection bias) leaving out crucial primary studies that may affect the results of the review.

In general, assessors found that AMSTAR-2 was more straightforward and user friendly than ROBIS.^{18 19} The two tools had similar inter-rater reliability.^{18 20 21} The range in time taken to use AMSTAR-2 was similar to ROBIS (14-60 v 16-60 min) across three comparison studies^{18 20 21} (table 1). ROBIS users required training and practice in using the tool^{22 23} and it was often understood and applied differently.²⁰ AMSTAR-2 has been criticised for unclear guidance on some items,²⁴⁻²⁶ which can lead to varying interpretations and applications. ROBIS is accompanied by voluminous guidance, which can be difficult to manage by the user.²¹⁻²³

While AMSTAR-2 and ROBIS are both widely used tools for assessing systematic reviews, in some situations, one may be preferred over the other. AMSTAR-2 may be preferred when:

- ▶ the primary focus is evaluating the methodological quality of a systematic review of interventions;

- ▶ the aim is to broadly assess aspects of review conduct, reporting comprehensiveness, and risk of bias; or
- ▶ a relatively quick and easy to use tool is sought, because AMSTAR-2 has fewer items compared with ROBIS.

ROBIS may be preferred when:

- ▶ the aim is to identify concerns with the review conduct that may point to risk of biases in the results and conclusions, as well as assessing relevance and minimising interpretation bias or 'spin';
- ▶ a more nuanced tool is sought, which may involve more thoughtful assessment and time, because ROBIS contains more items compared with AMSTAR-2;
- ▶ the aim is to assess multiple types of systematic reviews to compare risk of bias across them (eg, when preparing a clinical practice guideline).

Reporting and interpretation

When reporting and interpreting the overview results, assessors should note some key considerations with AMSTAR-2 and ROBIS assessments. Authors should first report methodological quality or bias assessment results by item, domain, and overall judgement. In addition, assessment should be reported at the outcome level as opposed to the systematic review level.¹⁸ Several responses to AMSTAR-2 item 13 (whether risk of bias was discussed or interpreted) are possible when multiple outcomes (eg, mortality and adverse events) are reported in one systematic review. Ideally, results of intervention overviews should be reported by qualifying the inherent methodological quality or risk of bias in the included systematic reviews as potential limitations.

Subgrouping systematic reviews by low and high risk of bias using ROBIS can be a great way to determine whether authors of reviews of interventions that have a high risk of bias over emphasised their findings and conclusions. Subgrouping also allows overview authors to exclude systematic reviews that are at a high risk of bias from the synthesis. However, using only one single criteria (ie, the systematic reviews at low risk of bias) for inclusion in analyses can result in unintended loss of information through exclusion of important systematic review data (eg, by excluding the systematic review with the greatest number of unique trials).

Conclusions

Overviews are used by guideline developers and policy makers to summarise large bodies of evidence in consideration of interventions of interest on a given topic. Using the appropriate tools to critically appraise included systematic reviews of intervention effects means that a complete assessment of methodological quality and all the potential biases are considered. Systematic reviews vary considerably by

method, how data are synthesised, and how results and conclusions are reported, therefore, an assessment of potential biases is necessary to consider their reproducibility, trustworthiness, and usefulness for end users. At this time, the recommended tools to assess methodological quality and bias among systematic reviews included in overviews are AMSTAR-2 and ROBIS. Proper use of these tools for this purpose requires training, time, and methodological insight.

AUTHOR AFFILIATIONS

¹Knowledge Translation Program, Li Ka Shing Knowledge Institute, St Michael's Hospital, Toronto, ON, Canada

²Cochrane Hypertension Review Group, Cochrane Canada, Vancouver, BC, Canada

³Ottawa Hospital, Ottawa, ON, Canada

⁴Ottawa Hospital Research Institute, Ottawa, ON, Canada

⁵School of Pharmaceutical Sciences University of Ottawa, Hôpital Montfort, Ottawa, ON, Canada

⁶Department of Hygiene, Social-Preventive Medicine and Medical Statistics, Aristotle University of Thessaloniki Faculty of Health Sciences, Thessaloniki, Central Macedonia, Greece

⁷Faculty of Health Sciences Brandenburg, Brandenburg Medical School Theodor Fontane Ruppiner Clinics, Neuruppin, Brandenburg, Germany

X Carole Lunny @carole_lunny, Pierre Thabet @PierreThabet and Konstantinos I Bougioukas @Bougioukas

Contributors CL conceived the idea of the study and drafted the manuscript; CL, PS, DW, KB, SK, and ABH edited the manuscript and read and approved the final manuscript.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests We have read and understood the BMJ policy on declaration of interests and declare the following interests: none.

Provenance and peer review Commissioned; externally peer reviewed.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iD

Carole Lunny <http://orcid.org/0000-0002-7825-6765>

REFERENCES

- Higgins JP, Thomas J, Chandler J, *et al.* Cochrane Handbook for Systematic Reviews of Interventions. John Wiley & Sons, 2019.
- Lunny C, Neelakant T, Chen A, *et al.* "Bibliometric study of 'Overviews of systematic reviews' of health interventions: evaluation of prevalence, citation and Journal impact factor". *Res Synth Methods* 2022;13:109–20. 10.1002/jrsm.1530
- Hartling L, Chisholm A, Thomson D, *et al.* A descriptive analysis of overviews of reviews published between 2000 and 2011. *PLoS One* 2012;7:e49667. 10.1371/journal.pone.0049667
- Hunt H, Pollock A, Campbell P, *et al.* An introduction to overviews of reviews: planning a relevant research question and objective for an overview. *Syst Rev* 2018;7:39:39. 10.1186/s13643-018-0695-8

- Pieper D, Buechter R, Jerinic P, *et al.* Overviews of reviews often have limited rigor: a systematic review. *J Clin Epidemiol* 2012;65:1267–73. 10.1016/j.jclinepi.2012.06.015
- Smith V, Devane D, Begley CM, *et al.* Methodology in conducting a systematic review of systematic reviews of healthcare interventions. *BMC Med Res Methodol* 2011;11:1–6. 10.1186/1471-2288-11-15
- Thomson D, Russell K, Becker L, *et al.* The evolution of a new publication type: steps and challenges of producing overviews of reviews. *Res Synth Methods* 2010;1:198–211. 10.1002/jrsm.30
- Lunny C, Brennan SE, McDonald S, *et al.* Toward a comprehensive evidence map of overview of systematic review methods: paper 2- risk of bias assessment; synthesis, presentation and summary of the findings; and assessment of the certainty of the evidence. *Syst Rev* 2018;7:159. 10.1186/s13643-018-0784-8
- Whiting P, Davies P, Savović J, *et al.* Chapter 5, phase 3: review of studies that have used the AMSTAR tool. evidence to inform the development of ROBIS, a new tool to assess the risk of bias in systematic reviews. 2013. Available: <http://www.robis-tool.info>
- Shea BJ, Reeves BC, Wells G, *et al.* AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of Healthcare interventions, or both. *BMJ* 2017;358:j4008. 10.1136/bmj.j4008
- Whiting P, Savović J, Higgins JPT, *et al.* ROBIS: a new tool to assess risk of bias in systematic reviews was developed. *J Clin Epidemiol* 2016;69:225–34. 10.1016/j.jclinepi.2015.06.005
- Lunny C, Veroniki A-A, Higgins JP. Methodological review of NMA bias concepts provides groundwork for the development of a list of concepts for potential inclusion in a new risk of bias tool for network meta-analysis (RoB NMA tool). *Researchsquare*; 2022. Available: <https://www.researchsquare.com/article/rs-2010358/v1>
- Lunny C, Veroniki AA, Hutton B, *et al.* Knowledge user survey and Delphi process to inform development of a new risk of bias tool to assess systematic reviews with network meta-analysis. *BMJ EBM* 2023;28:58–67. 10.1136/bmjebm-2022-111944
- Pollock M, Fernandes RM, Becker LA, *et al.* Chapter V: overviews of reviews. *Cochrane Handbook for Systematic Reviews of Interventions* Version 2020;6.
- Aromataris E, Fernandez RS, Godfrey C, *et al.* Methodology for JBI umbrella reviews. 2014.
- Normansell R, Sayer B, Waterson S, *et al.* Antibiotics for exacerbations of asthma. *Cochrane Database Syst Rev* 2018;6:CD002741. 10.1002/14651858.CD002741.pub2
- Puig L, Thom H, Mollon P, *et al.* Clear or almost clear skin improves the quality of life in patients with Moderate-To-Severe psoriasis: a systematic review and Meta-Analysis. *Acad Dermatol Venereol* 2017;31:213–20. 10.1111/jdv.14007
- Perry R, Whitmarsh A, Leach V, *et al.* A comparison of two assessment tools used in overviews of systematic reviews: ROBIS versus AMSTAR-2. *Syst Rev* 2021;10:273. 10.1186/s13643-021-01819-x
- Pieper D, Mathes T, Eikermann M. Can AMSTAR also be applied to systematic reviews of non-randomized studies *BMC Res Notes* 2014;7:609. 10.1186/1756-0500-7-609
- Gates M, Gates A, Duarte G, *et al.* Quality and risk of bias appraisals of systematic reviews are inconsistent across reviewers and centers. *J Clin Epidemiol* 2020;125:9–15. 10.1016/j.jclinepi.2020.04.026
- Pieper D, Puljak L, González-Lorenzo M, *et al.* Minor differences were found between AMSTAR 2 and ROBIS in the assessment of systematic reviews including both randomized and Nonrandomized studies. *J Clin Epidemiol* 2019;108:26–33. 10.1016/j.jclinepi.2018.12.004
- Banzi R, Cinquini M, Gonzalez-Lorenzo M, *et al.* Quality assessment versus risk of bias in systematic reviews: AMSTAR and ROBIS had similar reliability but differed in their construct and applicability. *J Clin Epidemiol* 2018;99:24–32. 10.1016/j.jclinepi.2018.02.024
- Bühn S, Mathes T, Prengel P, *et al.* The risk of bias in systematic reviews tool showed fair Reliability and good construct validity. *J Clin Epidemiol* 2017;91:121–8. 10.1016/j.jclinepi.2017.06.019
- Burda BU, Holmer HK, Norris SL. Limitations of A measurement tool to assess systematic reviews (AMSTAR) and suggestions for improvement. *Syst Rev* 2016;5:58. 10.1186/s13643-016-0237-1
- Faggion CM. Critical appraisal of AMSTAR: challenges, limitations, and potential solutions from the perspective of an assessor. *BMC Med Res Methodol* 2015;15:63. 10.1186/s12874-015-0062-6
- Wegewitz U, Weikert B, Fishta A, *et al.* Resuming the discussion of AMSTAR: what can (should) be made better. *BMC Med Res Methodol* 2016;16:111. 10.1186/s12874-016-0183-6